#### **ORIGINAL PAPER**



# Accuracy of diagnostic classification and clinical utility assessment of ICD-11 compared to ICD-10 in 10 mental disorders: findings from a web-based field study

Wolfgang Gaebel<sup>1,2</sup> · Johannes Stricker<sup>1,2</sup> · Mathias Riesbeck<sup>1</sup> · Jürgen Zielasek<sup>3</sup> · Ariane Kerst<sup>1,2</sup> · Eva Meisenzahl-Lechner<sup>1,2,3</sup> · Volker Köllner<sup>4,5</sup> · Matthias Rose<sup>5</sup> · Tobias Hofmann<sup>5</sup> · Ingo Schäfer<sup>6</sup> · Annett Lotzin<sup>6</sup> · Peer Briken<sup>7</sup> · Verena Klein<sup>7</sup> · Franziska Brunner<sup>7</sup> · Jared W. Keeley<sup>8</sup> · Julia Brechbiel<sup>8</sup> · Tahilia J. Rebello<sup>9</sup> · Howard F. Andrews<sup>9</sup> · Geoffrey M. Reed<sup>9,10</sup> · Ulrich Vogel<sup>11</sup> · Alkomiet Hasan<sup>12</sup> · Peter Falkai<sup>12</sup>

Received: 26 August 2019 / Accepted: 9 October 2019 © Springer-Verlag GmbH Germany, part of Springer Nature 2019

#### Abstract

In this web-based field study, we compared the diagnostic accuracy and clinical utility of 10 selected mental disorders between the ICD-11 Clinical Descriptions and Diagnostic Guidelines (CDDG) and the ICD-10 CDDG using vignettes in a sample of 928 health professionals from all WHO regions. On average, the ICD-11 CDDG displayed significantly higher diagnostic accuracy (71.9% for ICD-11, 53.2% for ICD-10), higher ease of use, better goodness of fit, higher clarity, and lower time required for diagnosis compared to the ICD-10 CDDG. The advantages of the ICD-11 CDDG were largely limited to new diagnoses in ICD-11. After limiting analyses to diagnoses existing in ICD-11 and ICD-10, the ICD-11 CDDG were only superior in ease of use. The ICD-11 CDDG were not inferior in diagnostic accuracy or clinical utility compared to the ICD-10 CDDG for any of the vignettes. Diagnostic accuracy was consistent across WHO regions and independent of participants' clinical experience. There were no differences between medical doctors and psychologists in diagnostic accuracy, but members of other health professions had greater difficulties in determining correct diagnoses based on the ICD-11 CDDG. In sum, there were no differences in diagnostic accuracy for diagnoses existing in ICD-10 and ICD-11, but the introduction of new diagnoses in ICD-11 CDDG give reason to expect a positive evaluation by health professionals in the implementation phase of ICD-11. Yet, training in ICD-11 is needed to further enhance the diagnostic accuracy.

**Keywords** International Classification of Diseases  $\cdot$  ICD-11  $\cdot$  Mental and behavioural disorders  $\cdot$  Field study  $\cdot$  Diagnosis  $\cdot$  Diagnostic accuracy  $\cdot$  Clinical utility

### Introduction

In May 2019, the 72nd World Health Assembly endorsed the Eleventh Revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-11). According to the World Health Organization (WHO), mental and behavioural disorders are among the top ranked

Wolfgang Gaebel wolfgang.gaebel@uni-duesseldorf.de

Extended author information available on the last page of the article

disorders contributing to the most years of life lost due to disability (YLDs) [1]. By 2030, mental disorders will be the "top drivers for lost output" in non-communicable diseases, putting them ahead of cardiovascular disorders [2]. Insufficient treatment and care of mental disorders is the main factor for the high indirect treatment costs, fourfold higher than the direct treatment costs; and lacking or incorrect diagnoses significantly contribute to the aforementioned deficits in treatment and care [3]. Thus, the diagnostic classification accuracy and clinical utility of ICD-11 will be a central prerequisite for the initiation of evidence-based and cost-effective treatment worldwide.

The majority of mental health professionals often or routinely consult diagnostic guidelines when determining initial diagnoses for their patients [4]. In these diagnostic

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s00406-019-01076-z) contains supplementary material, which is available to authorized users.

guidelines, there were significant changes from ICD-10 to ICD-11 [5]. Changes include modifications of the diagnostic criteria of well-established diagnoses (e.g., schizophrenia) and the inclusion of a set of new diagnoses (e.g., Complex Post-Traumatic Stress Disorder and Binge Eating Disorder). The ICD-11 Clinical Descriptions and Diagnostic Guidelines (CDDG) were developed to provide mental health professionals with clear diagnostic guidance. To test whether this objective was achieved, the first aim of this study was to compare clinicians' diagnostic accuracy (i.e., whether health professionals derive the correct diagnosis when applying diagnostic guidelines [6]) between the ICD-11 CDDG and the ICD-10 CDDG, which serve as a benchmark.

Besides diagnostic accuracy, clinical utility is highly relevant for diagnostic classification systems. A diagnostic classification system not perceived as clinically useful will not be implemented accurately in clinical practice [7, 8]. Consequently, increasing clinical utility was a primary aim of the ICD-11 revision process [8-10]. The WHO definition of clinically useful classification systems/constructs include their value in communication (e.g., between practitioners), implementation characteristics (i.e., ease of use, goodness of fit, clarity, time required for diagnosis) and their ability to aid in clinical decision-making to select appropriate interventions [10]. Therefore, to test acceptability of the guidelines, health professionals' perspectives and user experiences are of central interest in the evaluation of ICD-11. Thus, the second aim of this study was to compare the clinical utility of the ICD-11 CDDG and the ICD-10 CDDG.

Despite a number of published evaluative ICD-11 field studies [6, 11–14], factors influencing practitioners' diagnostic accuracy have only recently received empirical attention [15]. There is a possibility that characteristics of health professionals influence diagnostic accuracy. For example, health professionals with long clinical experience might face greater difficulty in determining ICD-11 diagnoses because they are more strongly accustomed to ICD-10. There might also be variation between different clinical professions in the ability to apply the ICD-11 CDDG or between health professionals from different WHO regions. Identifying these demographic influences will be crucial for target-oriented and well-tailored training programs in the implementation stage of ICD-11.

Perceived characteristics of the diagnostic guidelines themselves (e.g., clarity) may also influence diagnostic accuracy. Identifying characteristics of the ICD-11 CDDG that are associated with diagnostic accuracy can inform future revisions of diagnostic guidelines. Hence, the third aim of this study was to identify user-related variables (years of professional experience, profession, gender, WHO region) and clinical utility variables (ease of use, goodness of fit, clarity, time required for diagnosis) that are associated with clinical professionals' diagnostic accuracy in ICD-11 diagnoses.

To address the three aforementioned research aims, a randomized vignette-based field study was conducted by five German medical associations (German Society of Psychiatry, Psychotherapy and Psychosomatics, German Society for Psychosomatic Medicine and Medical Psychotherapy, German Society for Clinical Psychotherapy and Psychosomatic Rehabilitation, German Society for Psychotraumatology, and German Society for Sex Research) in coordination with the German Institute of Medical Documentation and Information. In this study, we investigated the diagnostic accuracy and perceived clinical utility of the ICD-11 CDDG and the ICD-10 CDDG for ten mental disorders.

#### Methods

#### Study procedure and design

This field study was conducted on the Global Clinical Practice Network (GCPN) platform, an international web-based network of mental health and primary care professionals, established by the World Health Organization's Department of Mental Health and Substance Abuse. Any mental health or primary care professionals, who completed clinical training, qualifying them to practice in the mental health field in their country, are eligible to register in the GCPN. Study procedures (see Fig. 1) were programmed with Qualtrics, a web-based survey software. All study materials were in



Fig. 1 Study procedure

English. Expert panels designed case vignettes reflecting ICD-11 diagnoses based on the guidelines for ICD-11 field studies [16] and determined correct diagnoses in a consensus process. Case vignettes for this study were selected by expert consensus taking into consideration the prevalence of the respective mental disorder and the degree of modification between ICD-11 and ICD-10. Overall, there were ten case vignettes: Schizophrenia, Schizoaffective Disorder, Bipolar Disorder Type II, Recurrent Depressive Disorder, Moderate Personality Disorder, Adjustment Disorder, Complex Post-Traumatic Stress Disorder (PTSD), Binge Eating Disorder, Bodily Distress Disorder, and Compulsive Sexual Behaviour Disorder. After receiving general study information and agreeing to participate, participants were sent the participation link. On clicking the participation link, participants were randomly assigned to a classification system (ICD-11 vs ICD-10), then randomly assigned to one of the ten vignettes (see Fig. 1). In the first step, participants were shown the ICD-11 or ICD-10 CDDG descriptions for the correct and several alternative diagnoses. Table S1 displays the correct diagnoses for the vignettes according to ICD-11 and ICD-10 as determined by the expert panels. After presentation of the CDDG, participants were presented with one randomly allocated case vignette that contained the description of a patient presenting with various symptoms. Participants were then given the choice to select a diagnosis (single choice) from a list of available ICD-11 or ICD-10 diagnoses. It was also possible to select "a different diagnosis" whereby participants were prompted to name another diagnosis or to select "no diagnosis". Throughout the survey, participants had the option to view the diagnostic guidelines, the case vignette and the list of diagnoses based on either one of the versions of ICD simultaneously. After selecting a diagnosis, participants rated the perceived clinical utility (ease of use, goodness of fit, clarity) of the diagnostic guidelines. In addition, the time required to come to a final diagnosis was documented. We used a between-subject design to avoid learning and carryover effects. The vignette approach allows for random allocation of participants to a diagnostic guideline (ICD-11 vs ICD-10) while keeping the clinical presentations (i.e., described symptoms) constant. Hence, we were able to isolate the effects of clinical guidelines on diagnostic accuracy and clinical utility. Other advantages of vignette studies include their ability to minimize assessment bias (e.g., observer effects [16]) and a high resemblance between clinicians' responses to vignettes and their responses to real-life situations [17].

An earlier smaller version of this dataset based on German-speaking participants only (n=319) was published previously to contribute to the feedback of the German Ministry of Health to the WHO regarding revision of the ICD-11 betadraft [11]. In this sample, we found a higher diagnostic accuracy for ICD-11 compared to ICD-10. However, time constraints and a low response rate lead to a relatively small sample size resulting in low statistical power for comparisons on the level of individual vignettes. Thus, we opened participation for all members of the GCPN and we extended the participation period leading to the inclusion of additional health professionals from all WHO regions in the present study (*n* overall = 928). The earlier publication additionally reported findings from a study that compared the consistency of diagnostic code assignment between ICD-11 and ICD-10 in a different sample (n = 120 German health professionals).

#### Participants

The study was conducted in a two-step convenience sampling procedure. In the first step, only German-speaking participants were included to contribute to a scheduled feedback of the German Ministry of Health to the WHO regarding revision of the ICD-11 betadraft [11]. These participants were recruited via email invitations to all members of the participating medical associations, the German association of psychotherapists, and of different project partners at cooperating mental health facilities (in total about 15,000 mental health professionals). Participants were invited to register at the GCPN platform and to subsequently participate in the study. In the second step, all health professionals already registered at the GCPN platform were invited to participate in the study.

Overall, 928 (465 female, 463 male) health professionals participated in this study. The mean age was 50.7 (11.4) years and the mean years of professional experience were 17.1 (10.6). 55.4% of the participants were from the European WHO Region (34.3% from Germany), followed by the Region of the Americas (22.1%), the South-East Asian Region (7.7%), the Western Pacific Region (6.9%), the African Region (5.9%), and the Eastern Mediterranean Region (2.0%). Of the participants, 47.1% indicated medicine as their profession (of which 84.9% indicated psychiatry as their medical discipline), 34.8% indicated psychology as their profession, and 18.1% indicated other health professions (4.6% counselling, 4.5% social work, 2.8% nursing, 2.5% occupational therapy, 0.6% sex therapy, 0.5% certified peer support worker, and 2.5% other). There were no significant differences between the two experimental groups (ICD-11 vs ICD-10) in demographic characteristics (all  $ps \ge 0.150$ , see Table S2 for details).

Between participants recruited in the first and second recruitment steps, there were no differences in diagnostic accuracy (p = 0.408), ease of use (p = 0.348), goodness of fit (p = 0.694), time required for diagnosis (p = 0.737), age (p = 0.078), and gender (p = 0.173). Participants from the first recruitment step rated the diagnostic guidelines as clearer compared to participants from the second recruitment step (p = 0.035) and had fewer years of professional experience (14.1 years compared to 18.5 years, p = 0.013). Also, a larger percentage of participants recruited in the second recruitment step worked in health professions other than medicine or psychology compared to participants recruited in the second recruitment step (recruitment step 1: 52.4% medicine, 44.9% psychology, 2.7% other health professions; recruitment step 2: 44.6% medicine, 30.1% psychology, 25.3% other health professions; p < 0.001).

#### Measures

#### **Diagnostic accuracy**

We assessed diagnostic accuracy as a dichotomous outcome (correct vs incorrect). Agreement of the participant's diagnostic decision with the diagnosis of the expert panel for the respective case vignette indicated a correct diagnosis. All other diagnostic decisions indicated an incorrect diagnosis.

#### **Clinical utility**

Three indicators were used to assess perceived clinical utility, each measured on a 4-point Likert scale: ease of use ('not at all easy to use' to 'extremely easy to use'), goodness of fit ('not at all accurate' to 'extremely accurate'), and clarity ('not at all clear' to 'extremely clear'). Additionally, the time interval from the presentation of the CDDG and the case vignette to the final selection of a diagnosis was recorded (i.e., the time required to come to a diagnostic decision).

#### Participants' characteristics

Participants indicated their age, years of professional experience, medical profession, and country of residence upon registration at the GCPN.

#### **Statistical analyses**

To compare diagnostic accuracy between ICD-11 and ICD-10, we conducted binary logistic regression analyses. In the next step, we compared assessments of clinical utility (i.e., ease of use, goodness of fit, clarity, time required for diagnosis) between ICD-11 and ICD-10. Since these variables were not normally distributed, we used the Mann–Whitney U test as a non-parametric test for independent samples. In the final step, we conducted two logistic regression analyses to test whether user-related variables (years of professional experience, profession, gender, WHO region) and clinical utility variables (i.e., ease of use, goodness of fit, clarity, time required for diagnosis) are associated with the diagnostic accuracy of ICD-11 diagnoses (correct vs not correct). Diagnostic accuracy of

ICD-11 diagnoses was regressed on years of professional experience, profession, gender, and WHO region (userrelated variables) in the first logistic regression and on ease of use, goodness of fit, clarity, and time required for diagnosis (clinical utility variables) in the second logistic regression. All analyses were conducted on the level of diagnostic classification systems (ICD-11 vs ICD-10; aggregated across vignettes) and on the level of individual vignettes, except for the analyses of variables predictive for diagnostic accuracy of ICD-11. We did not test the predictive effects of user-related and clinical utility variables on the accuracy of ICD-11 diagnoses on vignettelevel, because the number of participants per vignette was too low to estimate reliable values for these analyses. In addition, given the small number of participants in most countries, country data were aggregated on the level of WHO regions.

#### Results

#### **Diagnostic accuracy**

Table 1 displays the percentages of correct diagnoses for ICD-11 and ICD-10 and the results from the logistic regression analyses. Overall, diagnostic accuracy was higher for ICD-11 (correct diagnosis: 71.9%) compared to ICD-10 (correct diagnosis: 53.2%;  $\chi^2$  (1)=33.9, p<0.001). The odds for a correct diagnosis were 2.2 times higher for participants that gave a diagnosis based on ICD-11 compared to participants that gave a diagnosis based on ICD-10. Regarding specific vignettes, the diagnostic accuracy of ICD-11 was significantly higher for the Bipolar Disorder Type II vignette  $(\chi^2 (1) = 26.5, p < 0.001)$ , the Complex PTSD vignette  $(\chi^2 (1) = 26.5, p < 0.001)$ (1) = 13.7, p < 0.001), the Bodily Distress Disorder vignette  $(\chi^2(1)=20.3, p<0.001)$ , and the Compulsive Sexual Behaviour Disorder vignette ( $\chi^2$  (1) = 16.3, p < 0.001) compared to ICD-10. There was no vignette in which the diagnostic accuracy of participants in the ICD-11 group was significantly lower than the diagnostic accuracy of participant in the ICD-10 group. We additionally tested whether the diagnostic accuracy of ICD-11 was still higher compared to ICD-10 after excluding all vignettes that pertained to new diagnoses in ICD-11 (i.e., Bipolar Disorder Type II, Moderate Personality Disorder, Complex PTSD, Binge Eating Disorder, Bodily Distress Disorder, and Compulsive Sexual Behaviour Disorder). After limiting the analysis to diagnoses existing in both ICD-11 and ICD-10 (i.e., Schizophrenia, Schizoaffective Disorder, Recurrent Depressive Disorder, and Adjustment Disorder), there was no significant difference between the two classification systems ( $\chi^2$  (1)=0.0, p = 0.894).

Table 1Percentages of correctdiagnoses and results of thelogistic regression analysescomparing the diagnosticaccuracy of ICD-11 and ICD-10

Vignette	п	ICD-11 % correct	ICD-10 % correct	р	OR
Overall	928	71.9	53.2	< 0.001	2.2
Schizophrenia	94	74.4	78.4	0.647	0.8
Schizoaffective Disorder	95	63.5	44.2	0.062	2.2
Bipolar Disorder Type II	90	68.4	9.6	< 0.001	20.4
Recurrent Depressive Disorder	97	81.6	66.7	0.096	2.2
Moderate Personality Disorder	89	57.4	73.8	0.108	0.5
Adjustment Disorder	92	34.6	55.0	0.052	0.4
Complex PTSD	95	71.1	32.0	< 0.001	5.2
Binge Eating Disorder	92	86.5	87.5	0.892	1.1
Bodily Distress Disorder	89	95.5	37.8	< 0.001	34.6
Compulsive Sexual Behaviour Disorder	95	89.3	48.7	< 0.001	8.8

Statistically significant differences (p < 0.05) are highlighted in bold *OR* odds ratio

#### **Clinical utility**

A large majority of participants perceived the clinical utility of the ICD-11 CDDG as positive. 88.6% rated the ICD-11 CDDG as quite to extremely easy to use, 89.3% as quite to extremely accurate and 86.4% as quite to extremely clear. Table 2 displays the mean levels of ease of use, goodness of fit, clarity, and time required for diagnosis as well as the results of the Mann–Whitney U tests comparing these indicators of clinical utility between ICD-11 and ICD-10. We excluded participants as outliers that required an abnormally long or short time period for diagnosis (i.e.,  $\pm 2$  standard deviations) from the analysis of time required for diagnosis (n = 18). The ICD-11 CDDG were superior to the ICD-10 CDDG in all indicators of clinical utility. The ICD-11 CDDG displayed a higher perceived ease of use (p < 0.001), greater perceived goodness of fit (p < 0.001), greater perceived clarity (p < 0.001) and lower time required for diagnosis (p=0.007) compared to the ICD-10 CDDG. On the level of specific vignettes, the clinical utility of the ICD-11 CDDG was superior to the ICD-10 CDDG for the Binge Eating Disorder vignette, the Bodily Distress Disorder vignette, the Compulsive Sexual Behaviour Disorder vignette, and, to a lesser extent, for the Schizoaffective Disorder vignette.

We additionally tested whether the indicators of clinical utility were more favourable for ICD-11 compared to ICD-10 after excluding all vignettes that pertained to new diagnoses in ICD-11 (i.e., Bipolar Disorder Type II, Moderate Personality Disorder, Complex PTSD, Binge Eating Disorder, Bodily Distress Disorder, and Compulsive Sexual Behaviour Disorder). After limiting the analysis to diagnoses existing in ICD-11 and ICD-10 (i.e., Schizophrenia, Schizoaffective Disorder, Recurrent Depressive Disorder, and Adjustment Disorder), ease of use was rated as higher for ICD-11 compared to ICD-10 (p = 0.048), but there were no significant

differences in goodness of fit (p=0.080), clarity (p=0.082) and time required for diagnosis (p=0.752). Table S3 displays the clinical utility ratings by response category. There were no significant differences in the perceived clinical utility of the new ICD-11 diagnoses compared to the ICD-11 diagnoses that already existed in ICD-10 (p=0.233 for ease of use, p=0.762 for goodness of fit, and p=0.209 for clarity). Time required for diagnosis was lower for the new ICD-11 diagnoses (1.9 min) compared to the ICD-11 diagnoses that already existed in ICD-10 (2.6 min, p=0.003).

# Variables associated with the diagnostic accuracy of ICD-11 diagnoses

Table 3 displays the results from the logistic regression analysis in which the diagnostic accuracy of ICD-11 diagnoses was regressed on user-related variables (years of professional experience, profession, gender, WHO region). Diagnostic accuracy of ICD-11 diagnoses was independent of participants' years of professional experience (p = 0.961). There was no significant difference in the diagnostic accuracy between medical doctors (75.0% correct) and psychologists (71.4% correct, p = 0.117), but participants from other health professions (64.7% correct) had greater difficulty in determining correct diagnoses compared to medical doctors (p=0.012). There were no gender differences (p=0.387). Across WHO regions, diagnostic accuracy was relatively consistent. The diagnostic accuracy of participants from the European region (72.3% correct) was not statistically different from the Region of the Americas (74.5% correct), the South-East Asia Region (77.1% correct), the Western Pacific Region (71.9% correct), the African Region (54.5% correct), and the Eastern Mediterranean Region, (72.7% correct, all  $ps \ge 0.095$ ).

Vignette	Ease of use	(h		Goodness	of fit		Clarity			Time for d	agnosis (min	~
	ICD-11	ICD-10	d	ICD-11	ICD-10	d	ICD-11	ICD-10	р	ICD-11	ICD-10	d
Overall	3.2 (0.6)	3.0 (0.7)	< 0.001	3.2 (0.6)	3.0 (0.7)	< 0.001	3.1 (0.7)	3.0 (0.7)	< 0.001	2.2 (2.4)	2.6 (2.9)	0.007
Schizophrenia	3.1(0.6)	3.0 (0.7)	0.281	3.3 (0.5)	3.1 (0.6)	0.301	3.2 (0.6)	3.0 (0.7)	0.264	2.6 (3.0)	3.3 (4.0)	0.329
Schizoaffective Disorder	3.0 (0.7)	2.6 (0.9)	0.021	3.0(0.8)	2.9 (0.7)	0.313	3.1 (0.7)	2.8 (0.8)	0.015	2.1 (1.7)	3.1 (3.2)	0.251
Bipolar Disorder Type II	2.9 (0.7)	3.0(0.8)	0.762	2.8 (0.6)	3.0(0.6)	0.063	2.8 (0.8)	3.0 (0.7)	0.201	3.2 (3.2)	2.4 (2.4)	0.097
Recurrent Depressive Disorder	3.4(0.6)	3.5 (0.7)	0.778	3.5 (0.6)	3.3 (0.8)	0.390	3.3 (0.7)	3.1 (0.8)	0.399	2.0 (2.2)	1.4(1.0)	0.417
Moderate Personality Disorder	3.0(0.5)	3.0(0.7)	0.920	2.9 (0.5)	2.8 (0.7)	0.297	3.0(0.6)	2.9 (0.6)	0.193	2.9 (2.8)	3.0(3.4)	0.773
Adjustment Disorder	3.0 (0.5)	2.8 (0.6)	0.121	2.9 (0.5)	2.9 (0.4)	0.300	2.8 (0.7)	2.9 (0.7)	0.544	3.8 (3.4)	3.3 (3.8)	0.191
Complex PTSD	3.3 (0.7)	3.2~(0.6)	0.425	3.2 (0.7)	3.1 (0.6)	0.260	3.2 (0.7)	3.1 (0.6)	0.167	1.6(1.4)	2.1 (1.8)	0.103
Binge Eating Disorder	3.5(0.6)	3.0(0.7)	0.001	3.6(0.6)	2.9 (0.7)	< 0.001	3.4(0.6)	2.9 (0.6)	< 0.001	1.5 (1.4)	2.6 (2.1)	0.001
<b>Bodily Distress Disorder</b>	3.3(0.6)	3.1(0.8)	0.156	3.4~(0.6)	2.9 (0.7)	0.001	3.3~(0.6)	3.0 (0.7)	0.045	1.3 (1.3)	2.9 (3.6)	< 0.001
Compulsive Sexual Behaviour Disorder	3.3 (0.7)	2.9 (0.6)	0.005	3.3 (0.6)	2.8 (0.6)	< 0.001	3.3 (0.7)	2.8 (0.7)	0.003	1.4 (1.2)	2.5 (2.6)	0.031

Deringer

European Archives of Psychiatry and Clinical Neuroscience

Table 4 displays the results from the logistic regression analysis in which the diagnostic accuracy of ICD-11 diagnoses was regressed on variables indicating clinical utility (ease of use, goodness of fit, clarity, time required for diagnosis). Goodness of fit was positively associated with overall diagnostic accuracy (p=0.018). Ease of use (p=0.774) and clarity (p=0.107) were not significantly associated with diagnostic accuracy after controlling for the other variables in the regression. Additionally, time for diagnosis was negatively related to diagnostic accuracy (p = 0.037) indicating that participants who required less time to come to a diagnostic decision had a higher diagnostic accuracy.

# Discussion

This international web-based field study provided insight into how well the ICD-11 CDDG function when applied by health professionals. Overall, the use of the ICD-11 CDDG led to an increased percentage of correctly selected diagnoses compared to the ICD-10 CDDG. Additionally, participants' experiences with the ICD-11 CDDG were rated as remarkably positive. The time required for diagnosis was lower and ratings of clinical utility were more favourable for the ICD-11 CDDG compared to the ICD-10 CDDG. However, on the level of specific diagnoses, advantages of the ICD-11 CDDG over the ICD-10 CDDG were identified only for some diagnoses and were largely limited to new diagnoses in ICD-11.

The on average superior diagnostic accuracy of the ICD-11 CDDG compared to the ICD-10 CDDG is in line with a previous ecological field study that identified superior reliabilities of the ICD-11 CDDG compared to previously reported reliabilities of the ICD-10 CDDG [13]. Taken together, these findings might be interpreted as a preliminary indicator for positive outcomes of the ICD-revision process. However, we found no significance difference in diagnostic accuracy for diagnoses existing in ICD-11 and ICD-10. Thus, the advantage of the ICD-11 over the ICD-10 was mainly driven by the inclusion of new diagnoses. Additionally, the diagnostic accuracy of ICD-11 diagnoses in this study should be improvable (71.9%). This might be explained by the complexity of the case vignettes or by the possibly insufficient training of health professionals in the ICD-11 guidelines.

Additionally, on the level of specific diagnoses, diagnostic accuracy was substantially lower for some ICD-11 diagnoses (e.g., 34.6% for Adjustment Disorder) compared to others (e.g., 95.5% for Bodily Distress Disorder). Adjustment Disorder is one of the most frequently used diagnoses [18]. ICD-10 defines Adjustment Disorder as a reaction to an identifiable stressor characterized by a wide range of impairments of social or occupational functioning and **Table 3** Results of the logistic regression analysis testing the associations between userrelated variables and diagnostic accuracy of the ICD-11 guidelines (n = 466)

Predictor	β	SE	$\chi^2$	df	р	OR
Clinical experience	0.0	0.0	0.0	1	0.961	1.0
Profession						
Medicine	Ref.					
Psychology	- 0.4	0.3	2.5	1	0.117	0.7
Other health professions	- 0.9	0.4	6.2	1	0.012	0.4
Gender	0.2	0.2	0.7	1	0.387	1.2
WHO region						
European Region	Ref.					
Region of the Americas	0.6	0.3	2.7	1	0.099	1.7
South–East Asia Region	0.3	0.4	0.4	1	0.526	1.3
Western Pacific Region	0.1	0.4	0.1	1	0.743	1.2
African Region	- 0.7	0.4	2.8	1	0.095	0.5
Eastern Mediterranean Region	0.0	0.7	0.0	1	0.953	1.0

Statistically significant difference (p < 0.05) is highlighted in bold

Gender: 1 male, 2 female

SE standard error, df degrees of freedom, OR odds ratio, Ref. reference category

**Table 4** Results of the logisticregression analysis testing theassociations between perceivedclinical utility and diagnosticaccuracy of the ICD-11guidelines (n=428)

Predictor	β	SE	$\chi^2$	df	р	OR
Ease of use	- 0.1	0.3	0.1	1	0.774	0.9
Goodness of fit	0.6	0.3	5.6	1	0.018	1.9
Clarity	0.4	0.2	2.6	1	0.107	1.5
Time for diagnosis	- 0.1	0.0	4.3	1	0.037	0.9

Statistically significant differences (p < 0.05) are highlighted in bold

SE standard error, df degrees of freedom, OR odds ratio

symptoms (e.g., depression, anxiety), none of which are unique to Adjustment Disorder. Adjustment Disorder was subject to substantive change from ICD-10 to ICD-11 and is now defined as comprising two symptoms: preoccupation with the stressor or its consequences and failure to adapt to the stressor (for a more detailed discussion of Adjustment Disorder in ICD-11, see [19]). Our results showed that the shift from a broad residual category to a more clearly defined diagnosis might have led to diagnostic difficulties. Hence, the effectiveness of trainings on ICD-11 classification might be enhanced by specifically highlighting changes in the CDDG of diagnoses whose labels remained despite substantive changes in the classification criteria (e.g., Adjustment Disorder).

Similar to DSM-5 [20], improving clinical utility was a central goal in the development of ICD-11 [8–10]. The finding that all four indicators of clinical utility (ease of use, goodness of fit, clarity, and time required for diagnoses) were more favourable for the ICD-11 CDDG compared to the ICD-10 CDDG provides preliminary evidence that the focus on clinical utility in the ICD-11 revision process was met by some success. On the level of specific diagnoses, comparisons with the ICD-10 CDDG regarding clinical utility were particularly favourable for three new diagnoses (i.e., Binge Eating Disorder, Bodily Distress Disorder, and Compulsive Sexual Behaviour Disorder) and, to a lesser extent, for Schizoaffective Disorder. On the descriptive level, the diagnostic accuracy for Schizoaffective Disorder was also higher for ICD-11 (63.5%) compared to ICD-10 (44.2%), although this comparison did not reach statistical significance (p = 0.062).

Since the vignettes were designed based on ICD-11 diagnoses, improved diagnostic accuracy and clinical utility of new ICD-11 diagnoses over ICD-10 diagnoses may not seem like a surprising finding. Yet, although ICD-10 has been used for decades, participants from the ICD-10 group did not display a higher diagnostic accuracy for any of the vignettes compared to participants from the ICD-11 group. Thus, implementation of the ICD-11 will not lead to a drop in diagnostic accuracy, which might have been expected based on mental health professionals' limited experience with and training on the ICD-11. Additionally, there was no indication of inferior clinical utility of the ICD-11 CDDG compared to the ICD-10 CDDG for any of the vignettes. Also for diagnoses existing in ICD-11 and in ICD-10, we found a superior ease of use for the ICD-11 CDDG compared to the ICD-10 CDDG. Thus, overall, the favourable clinical utility ratings of the ICD-11 CDDG in this study give reason to expect a positive evaluation by health professionals in the implementation phase.

Diagnostic accuracy was largely consistent across WHO regions. This finding provides some support for the global applicability of the ICD-11 CDDG. Improvements in diagnostic accuracy do not seem to be limited to specific groups of individuals, which might result from the international and collaborative nature of the revision process. Yet, health professionals from the African region (54.5% correct) descriptively seemed to have somewhat greater difficulties determining diagnoses based on the ICD-11 CDDG compared to other WHO regions (ranging from 72.3 to 77.1% correct). This might be addressed by specific training programs and culturally adapted training materials. However, this finding should not be over-interpreted because of the lack of statistical significance and the low number of participants from the African region. Regarding health professions, participants that were not medical doctors or psychologists (e.g., social workers) had greater difficulties in determining correct diagnoses. Since the role of these professions in mental health treatment varies strongly between countries, there is a risk that non-medical and non-psychological professions might be overlooked in diagnostic trainings, which will have to be addressed in the implementation stage of ICD-11. Regarding clinical utility variables, only goodness of fit was uniquely related to the diagnostic accuracy of ICD-11 diagnoses. Future research is needed to investigate how this specific aspect of clinical utility may be enhanced.

#### Limitations

There are several noteworthy limitations to this study. First, there is some concern over the artificiality of vignette studies, which might not accurately reflect the complexity of real-life situations [21]. Nevertheless, vignette studies were shown to approximate behaviour in real-world settings [17] and provide a suitable means for the investigation of diagnostic accuracy and clinical utility of diagnostic guidelines [16]. The overall consistency with a previous large-scale ecological field study [13, 14] additionally supports the validity of the results from our vignette approach. Second, the majority of case vignettes (6 out of 10) were designed based on diagnoses that were newly introduced in ICD-11. Thus, a selection of different vignettes might have led to different results regarding the differences in accuracy and clinical utility between the ICD-11 CDDG and the ICD-10 CDDG. Third, despite comprising health professionals from all WHO regions, our convenience sample might not be representative of all mental health professionals globally. For example, members of the GCPN might be more positive about the introduction of ICD-11 compared to health professionals who did not register for the GCPN and, therefore, did not participate in this study.

# Conclusion

Despite these limitations, this study provides valuable insights into experiences of health professionals with the ICD-11 CDDG. Diagnostic accuracy was superior to the ICD-10 CDDG and the clinical utility of the ICD-11 CDDG was positively perceived. Yet, improvements seem to be largely limited to the newly introduced diagnoses. These findings add to the growing empirical basis for the worldwide introduction of ICD-11.

Acknowledgements The collection of data from German-speaking participants (recruitment step 1) was funded by the German Federal Ministry of Health, Grant no. D105–01\_09\_ICD-11\_TMF\_DGPPN. We thank the following institutions and persons for their support in conducting this study: DGPPN (Prof. Dr. Arno Deister, Julie Holzhausen, Gabriel Gerlinger), TMF (Dr. Sebastian Semler and Dr. Annette Pollex-Krüger), DIMDI (Dr. Stefanie Weber, Maria Lange), KKG (Prof. Dr. Holger Reinecke and Prof. Dr. Jürgen Stausberg), BDK (Prof. Dr. Thomas Pollmächer), BPtK (Dr. Dietrich Munz and Dr. Tina Wessels), BVDN (Dr. Sabine Köhler), BVDP (Dr. Christa Roth-Sackenheim).

#### **Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- World Health Organization (2018) Disease burden 2000–2016. https://www.who.int/healthinfo/global\_burden\_disease/GHE20 16\_YLD\_Global\_2000\_2016\_.xls?ua=1. Accessed 25 Jul 2019
- Bloom DE, Cafiero ET, Jané-Llopis E et al (2012) The global economic burden of non-communicable diseases. In: PGDA working papers 8712, program on the global demography of aging
- Mack S, Jacobi F, Gerschler A et al (2014) Self-reported utilization of mental health services in the adult German population– evidence for unmet needs? Results of the DEGS1-Mental Health Module (DEGS1-MH). Int J Methods Psychiatr Res 23(3):289– 303. https://doi.org/10.1002/mpr.1438
- First MB, Rebello TJ, Keeley JW et al (2018) Do mental health professionals use diagnostic classifications the way we think they do? A global survey. World Psychiatry 17(2):187–195. https://doi. org/10.1002/wps.20525
- Reed GM, First MB, Kogan CS et al (2019) Innovations and changes in the ICD-11 classification of mental, behavioural and neurodevelopmental disorders. World Psychiatry 18(1):3–19. https://doi.org/10.1002/wps.20611
- Peterson DL, Webb CA, Keeley JW et al (2019) The reliability and clinical utility of ICD-11 schizoaffective disorder: a field trial. Schizophr Res. https://doi.org/10.1016/j.schres.2019.02.011
- Keeley JW, Reed GM, Roberts MC et al (2016) Developing a science of clinical utility in diagnostic classification systems: field study strategies for ICD-11 mental and behavioral disorders. Am Psychol 71(1):3–16. https://doi.org/10.1037/a0039972

- Reed GM, Roberts MC, Keeley J et al (2013) Mental health professionals' natural taxonomies of mental disorders: implications for the clinical utility of the ICD-11 and the DSM-5. J Clin Psychol 69(12):1191–1212. https://doi.org/10.1002/jclp.22031
- International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders (2011) A conceptual framework for the revision of the ICD-10 classification of mental and behavioural disorders. World Psychiatry 10(2):86–92. https://doi. org/10.1002/j.2051-5545.2011.tb00022.x
- Reed GM (2010) Toward ICD-11: improving the clinical utility of WHO's International Classification of mental disorders. Prof Psychol Res Pract 41(6):457–464. https://doi.org/10.1037/a0021 701
- Gaebel W, Riesbeck M, Zielasek J et al (2018) Internetbasierte Untersuchungen zur diagnostischen Klassifikation und Kodierung psychischer Störungen im Vergleich von ICD-11 und ICD-10. Fortschritte Neurol Psychiatrie 86(03):163–171. https://doi. org/10.1055/s-0044-100508
- Medina-Mora ME, Robles R, Rebello TJ et al (2019) ICD-11 guidelines for psychotic, mood, anxiety and stress-related disorders in Mexico: clinical utility and reliability. Int J Clin Health Psychol 19(1):1–11. https://doi.org/10.1016/j.ijchp.2018.09.003
- Reed GM, Sharan P, Rebello TJ et al (2018) The ICD-11 developmental field study of reliability of diagnoses of high-burden mental disorders: results among adult patients in mental health settings of 13 countries. World Psychiatry 17(2):174–186. https ://doi.org/10.1002/wps.20524

- Reed GM, Keeley JW, Rebello TJ et al (2018) Clinical utility of ICD-11 diagnostic guidelines for high-burden mental disorders: results from mental health settings in 13 countries. World Psychiatry 17(3):306–315. https://doi.org/10.1002/wps.20581
- Brechbiel JK, Keeley JW (2019) Pathways linking clinician demographics to mental health diagnostic accuracy: an international perspective. J Clin Psychol. https://doi.org/10.1002/jclp.22804
- Evans SC, Roberts MC, Keeley JW et al (2015) Vignette methodologies for studying clinicians' decision-making: validity, utility, and application in ICD-11 field studies. Int J Clin Health Psychol 15(2):160–170. https://doi.org/10.1016/j.ijchp.2014.12.001
- Peabody JW, Luck J, Glassman P et al (2000) Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. J Am Med Assoc 283(13):1715–1722. https://doi.org/10.1001/ jama.283.13.1715
- Reed GM, Correia JM, Esparza P (2011) The WPA–WHO global survey of psychiatrists' attitudes towards mental disorders classification. World Psychiatry 10(2):118–131. https://doi. org/10.1002/j.2051-5545.2011.tb00034.x
- Kazlauskas E, Zelviene P, Lorenz L et al (2017) A scoping review of ICD-11 adjustment disorder research. Eur J Psychotraumatol. https://doi.org/10.1080/20008198.2017.1421819
- 20. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th edn. APA, Washington, DC
- 21. Cullen S (2010) Survey-driven romanticism. Rev Philos Psychol 1(2):275–296. https://doi.org/10.1007/s13164-009-0016-1

# Affiliations

Wolfgang Gaebel<sup>1,2</sup> · Johannes Stricker<sup>1,2</sup> · Mathias Riesbeck<sup>1</sup> · Jürgen Zielasek<sup>3</sup> · Ariane Kerst<sup>1,2</sup> · Eva Meisenzahl-Lechner<sup>1,2,3</sup> · Volker Köllner<sup>4,5</sup> · Matthias Rose<sup>5</sup> · Tobias Hofmann<sup>5</sup> · Ingo Schäfer<sup>6</sup> · Annett Lotzin<sup>6</sup> · Peer Briken<sup>7</sup> · Verena Klein<sup>7</sup> · Franziska Brunner<sup>7</sup> · Jared W. Keeley<sup>8</sup> · Julia Brechbiel<sup>8</sup> · Tahilia J. Rebello<sup>9</sup> · Howard F. Andrews<sup>9</sup> · Geoffrey M. Reed<sup>9,10</sup> · Ulrich Vogel<sup>11</sup> · Alkomiet Hasan<sup>12</sup> · Peter Falkai<sup>12</sup>

- Department of Psychiatry, Medical Faculty, LVR-Klinikum Düsseldorf, Heinrich-Heine-University, Bergische Landstr. 2, 40629 Düsseldorf, Germany
- <sup>2</sup> WHO Collaborating Centre for Quality Assurance and Empowerment in Mental Health, Düsseldorf, Germany
- <sup>3</sup> LVR-Institute for Healthcare Research, Cologne, Germany
- <sup>4</sup> Department of Psychosomatic Medicine, Rehabilitation Center Seehof, Federal German Pension Agency, Teltow, Berlin, Germany
- <sup>5</sup> Department of Psychosomatic Medicine, Center for Internal Medicine and Dermatology, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
- <sup>6</sup> Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

- <sup>7</sup> Institute for Sex Research, Sexual Medicine and Forensic Psychiatry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
- <sup>8</sup> Department of Psychology, Virginia Commonwealth University, Richmond, USA
- <sup>9</sup> Columbia University Global Mental Health Program, New York, USA
- <sup>10</sup> Department of Mental Health and Substance Abuse, WHO, Geneva, Switzerland
- <sup>11</sup> German Institute of Medical Documentation and Information, Cologne, Germany
- <sup>12</sup> Department of Psychiatry and Psychotherapy, University Hospital LMU Munich, Munich, Germany